LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2025.2463082

REGULAR ARTICLE

When is a ranana a banana? Disentangling the mechanisms of error repair and word learning

Laurel Brehm[a,b], Nora Kennis[a] and Christina Bergmann[a,c]

[a]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands; [b]Department of Linguistics, University of California, Santa Barbara, CA, USA; [c]Recruiting and Talent Management, Osnabrück University of Applied Sciences, Osnabrück, Germany

ABSTRACT
When faced with an ambiguous novel word such as ‘ranana’, how do listeners decide whether they heard a mispronunciation of a familiar target (‘banana’) or a label for an unfamiliar novel item? We examined this question by combining visual-world eye-tracking with an offline forced-choice judgment paradigm. In two studies, we show evidence that participants entertain repair and novel label interpretations of novel words that were created by editing a familiar target word in multiple phonetic features (Experiment 1) or a single phonetic feature (Experiment 2). Repair (‘ranana’ = a banana) and learning (‘ranana’ = a novel referent) were both common interpretation strategies, and learning was strongly associated with visual attention to the novel image after it was referred to in a sentence. This indicates that repair and learning are both valid strategies for understanding novel words that depend upon a set of similar mechanisms, and suggests that attention during listening is causally related to whether one learns or repairs.

ARTICLE HISTORY
Received 5 August 2024
Accepted 21 January 2025

KEYWORDS
Noisy-channel processing; word learning; visual-world paradigm; speech errors; fast-mapping

1. Introduction

Language in the world is full of variability: Speakers vary in their dialect, accent, and the success of their target utterance production, which can make the acoustic forms of words and utterances difficult for listeners to predict. However, listeners seem to be skilled at managing to comprehend a speaker’s intended message most of the time. Two domains of literature provide different explanations as to how listeners cope with variability. First, literature on *non-literal processing* (extracting a meaning that differs from the literal compositional meaning of an utterance, e.g. Christianson et al., 2001; Ferreira, 2003; Gibson et al., 2013; Levy, 2008) shows that listeners frequently repair errors or non-canonical utterances into a form that is more expected. In other words, listeners can understand novel forms by performing a repair: the novel form ‘ranana’ is close enough to the known word ‘banana’ that it can be understood to mean ‘banana’. At the same time, listeners can also quickly learn meanings for novel word forms. Literature on *fast-mapping* (e.g. Coutanche & Thompson-Schill, 2014; Halberda, 2006; Storkel et al., 2006; Trueswell et al., 2013) shows that even adult listeners can learn new word-referent pairings with as little as one exposure. This shows how listeners can also understand novel forms by learning to map them to a referent: the novel form ‘ranana’ might instead be a label for a novel object. In some situations it may be unclear which of these strategies – repair or learning – to undertake. For example, some novel lexical items are also plausible errors because of their similarity to preexisting words: consider, for example, low-frequency words such as ‘bract’ (similar to ‘brat’ and ‘tract’) or slang such as ‘stan’ (similar to ‘stand’ and ‘stain’). We ask how adult listeners decide which strategy to pursue. In two studies, we combine online and offline measures to show how listeners recover meaning from ambiguous novel labels that are plausible as either mispronunciations or new words. This sheds light on how people navigate language in realistically-uncertain contexts.

1.1. Repair versus learning

A body of literature on non-literal processing exists under the *noisy-channel* (e.g. Gibson et al., 2013; Levy, 2008; Levy et al., 2009) and the *good-enough processing* (e.g. Christianson et al., 2001; Ferreira, 2003) frameworks.

CONTACT Laurel Brehm laurel.brehm@gmail.com, lbrehm@ucsb.edu South Hall 3432, Santa Barbara, CA, 93111
Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/23273798.2025.2463082

Though there are differences in the underlying assumed mechanisms, both frameworks can be used to explain the results of the current study, which is why we group them together for present purposes.

The premise of noisy-channel theory is that errors tend to be corrected when relatively few edits will turn the input into a more likely utterance. This is evidenced by the fact that small errors are more likely to be repaired (Gibson et al., 2013; Norris & McQueen, 2008), and more likely to be ignored than large errors in an error-detection paradigm (Cole, 1973). In contrast, good-enough processing focuses mainly on semantically or syntactically unlikely but grammatical sentences, which are also often reinterpreted to a more plausible outcome; the framework suggests that this occurs because fast heuristics are used instead of a slower, more effortful literal processing (see e.g. Karimi & Ferreira, 2016 for review). However, both claim that non-literal processing occurs, and that it occurs because of expectations derived from past language experience.

Non-literal interpretation occurs in domains as diverse as thematic role assignment (Ferreira, 2003), subject-verb agreement (Brehm et al., 2019, 2021; Keshev & Meltzer-Asscher, 2021; Patson & Husband, 2016) syntactic blends (Brehm et al., 2019, 2021; Frazier & Clifton Jr, 2011, 2015), and semantic illusions (Sanford & Sturt, 2002). Similarly, a wide range of cues impact non-literal interpretation rates, including the surrounding sentence context (Christianson & Luke, 2011; Lowder & Ferreira, 2019), the presence of filled pauses (Bailey & Ferreira, 2003; Lowder & Ferreira, 2019), the type of errors made by the speaker (Ryskin et al., 2018) and the perceived unreliability of the speaker (Brehm et al., 2019; Gibson et al., 2013, 2017). This shows how comprehenders take an active role in understanding sentences, integrating a variety of cues from the environment to form expectations about the most likely utterance given its context. The implication is that where a comprehender encounters a novel form, contextual cues and the overall likelihood of the form itself may lead to a change in interpretation, in favour of a familiar plausible alternative.

However, novel words also continuously enter the lexicon, with vocabulary size increasing until old age (e.g. Verhaeghen, 2003). This means that comprehenders need to keep open the possibility that a novel form corresponds to a new word that they have yet to learn. Because fast mapping was first studied in children, we draw upon both adult and child literatures to motivate our predictions here. Evidence from both literatures shows that successful learning of a label paired with a novel referent can happen with as little as one exposure (*fast mapping* in children: Carey & Bartlett, 1978; Heibeck & Markman, 1987 and adults: Coutanche & Thompson-Schill, 2014, 2015). Importantly, this work shows that in cases when there is a novel item present in a scene or array, children (Spiegel & Halberda, 2011) and adults (Coutanche & Thompson-Schill, 2014; Halberda, 2006; Storkel et al., 2006; Trueswell et al., 2013) are able to map a novel word to the item without it being explicitly labelled.

We propose that repair and fast-mapping are *mutually exclusive strategies* in some situations. We test this question in an adult population in the current paper, but draw upon both first- and second-language acquisition literatures to motivate the predictions. Investigating this hypothesis gives us a deeper understanding of the relationship between repair and learning and whether the same mechanisms support each factor. A first piece of evidence for the hypothesis is that fast-mapping is fairly rare in both children and adults : even if we assume that children are more adept language learners than adults are Newport et al. (2001), they are still not likely to learn new items on one exposure. This is the case especially in overt recall tasks (where learning occurs for a minority of participants in Carey & Bartlett, 1978 and on a minority of trials in Coutanche & Thompson-Schill, 2014), but the effect is also weak in recognition tasks despite the task being arguably easier (occurring on about half of trials in Coutanche & Thompson-Schill, 2014). Second, developmental work directly supports the idea that repair and novel word learning are often in opposition: Children are much more likely to select a known item than a novel item as a referent to words that could be novel items or likely mispronunciations, such as 'fesh' or 'yeaf' in forced-choice, mouse-tracking, or eye-tracking tasks (Creel, 2012; Krueger et al., 2018; Swingley, 2016). This opens the question of whether the same pattern holds throughout life, such that adults are likely to perform similarly. Finally, further evidence for the hypothesis comes from an adult language learning literature where larger learning effects occur for word forms which are more distant from competitors already in the lexicon (Storkel et al., 2006). This literature shows that having strong evidence that an item is novel improves word learning in adults. The noisy channel hypothesis predicts that the reverse is also true– evidence that an item is mispronounced increases repairs– suggesting that the two strategies may trade off.

In order to test whether repair and fast-mapping are indeed in opposition, we made several methodological innovations. First, to show independent evidence for repair and fast-mapping, one needs to measure evidence for both strategies for the same items. To do this, we developed a novel judgment elicitation approach that can be used with a signal-detection analysis. Materials were presented to participants in two phases: a *learning phase*, where participants got exposure to novel words and pictures paired together in scenes. We then elicited judgments in a *test phase* by asking participants whether images and audio labels do or do not go together. These judgments were performed by each participant on combinations of both images and both labels, which means that for any item, we can observe evidence for repair, for learning, for both, or for neither. This is sketched out in Figure 1.

In our analysis, which is inspired by a binomial model implementation of signal-detection theory (DeCarlo, 1998), we use difference scores to estimate rates of repair and learning on items separately from each other and in light of false alarms, providing a clean measure of the underlying interpretation. By subtracting conditions from each other (e.g. B2 from B1; D2 from D1 in Figure 1) and by separately estimating false alarm rates (conditions C1 and C2) from hit rates (conditions A1 and A2), we can dissociate repair, from learning, from false alarm errors. In other words: if participants decide that 'ranaan' maps to the banana, they should respond 'yes' to A1 and A2, while if participants decide that 'ranaan' maps to the novel object, they should respond 'yes' to D1 and D2.

In our experiments, we asked participants to perform two judgments on each trial, so that we can measure what is interpreted as the referent to *each word*. Making multiple judgments on each test item also allows us to indirectly assess the independence of strategies. Since participants can freely answer 'yes' on both judgments, we can also use examine the *sum* of repair and learning: if participants are entertaining both strategies, the rate of repair (B minus D) and the rate of learning (D1 minus D2) should sum to more than one (assuming a low rate of false alarms), whereas if the two strategies are mutually exclusive, their sum should be about one.

We also put the mechanisms for learning and repair to the test by manipulating several properties that existing theories predict to be important. One of these is the availability of a plausible same category novel referent in a scene. On all trials, there were two items of the category that was referred to in the sentence (e.g. 'fruit'), one known and one novel. If there are two possible referents that match the category mentioned in the narrative, e.g. one familiar and one novel kind of fruit, disjunctive reasoning dictates that a novel label should apply to the novel referent and not the familiar one. This seems to be the best explanation for the fast-mapping observed in earlier studies (Halberda, 2006; Trueswell et al., 2013). In our study, we always offer a novel referent in the learning phase as one plausible target for novel labels out of three possible objects, and manipulate whether the narrative in a specific scene highlights the novel referent with a contextually-matching adjective like 'new'. We track participants' attention to these referents with an eye-tracker to directly link processing during the learning phase with performance in the test phase, shedding light on what patterns of processing at encoding lead to repair and to learning.

We also explore whether learning and repair effects are modulated by individual experience. This provides insight into the mechanisms that support learning and repair. Existing evidence suggests that skilled language users (individuals with large vocabularies and strong reading, writing, and listening skills) have higher-quality representations in their mental lexicon (Perfetti, 2007; Perfetti & Hart, 2002). This promotes further acquisition: Skilled-reading adults (Perfetti et al., 2005) and children with large vocabularies (Duff et al., 2015; Fernald et al., 2006; Gershkoff-Stowe & Hahn, 2007) acquire new words faster than individuals of various ages with weaker language skills. Foreign language experience may also have a similar effect, since experienced language learners (bilingual or multilingual individuals) more quickly acquire novel foreign words than inexperienced language learners in a variety of paradigms (Kaushanskaya & Marian, 2009a, 2009b; Papagno & Vallar, 1995; Van Hell & Mahn, 1997, though c.f. Muench & Creel, 2013). The implication is that a variety of individual differences in learning ability and experience may sway whether fast mapping or repair occurs.

1.2. Current study

When faced with an unknown word such as 'ranana' in the face of a plausible unknown referent, comprehenders have a choice to repair the word to something familiar ('banana'), or to interpret it as a novel label for a novel referent and then acquire that word into their lexicon via fast mapping. In order to shed light on the mechanisms contributing to learning and repair, we conducted a two phase experiment. In the *learning phase*, we presented individuals with a set of familiar and novel words set in an engagement-boosting storybook context, and then in the *test phase*, we examined how these words were interpreted differently depending on the learning phase context.

On each critical learning phase trial, there was a familiar consonant-initial target (e.g. a banana; 'banaan' in Dutch), a same-category novel image (e.g. a prickly-pear fruit unknown to our Dutch participants), and a familiar, different-category distractor image (e.g. Lego blocks). In Experiment 1, we used novel words with approximately three phonetic features changed ('ranaan') and in Experiment 2, we used novel words

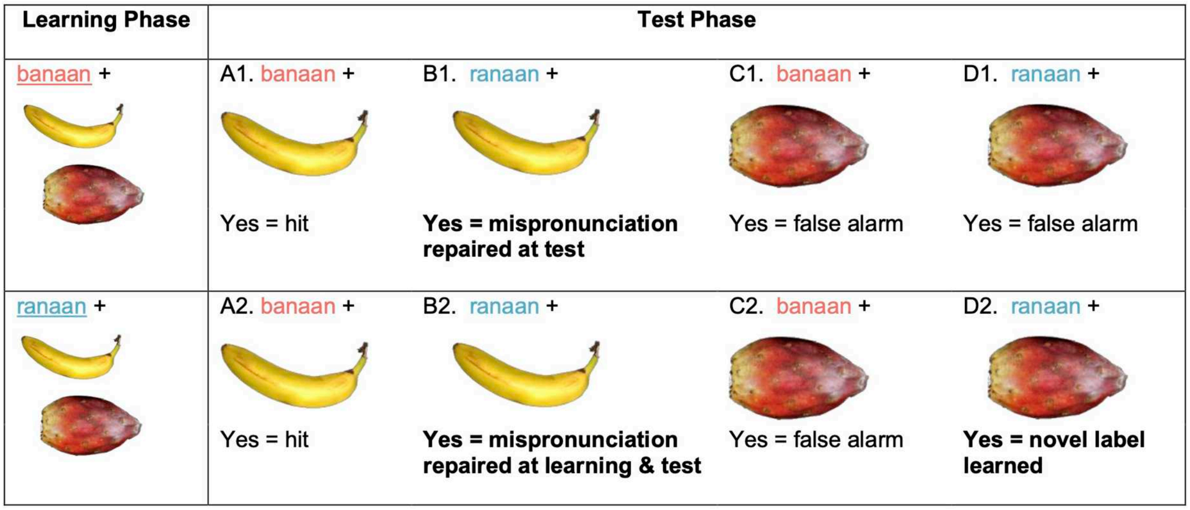

Figure 1. Inferences about participant 'yes' judgments in test phase trials conditional upon exposure in learning phase. Materials were presented in Dutch: 'banaan' is the Dutch word for 'banana'.

with approximately one feature changed ('panaan')[1]. These labels appeared after a contextually-constraining adjective that highlighted either the familiar item ('a familiar fruit') or a novel item ('a new fruit'), inspired by Halberda (2006) and Kukona et al. (2016).

We tracked people's eyes in the learning phase while they listened to the story. This gives us a fine-grained, temporally-sensitive measure of the way various parts of the sentence change attention (whether implicitly or explicitly directed) to the various referents (as in e.g. Altmann & Kamide, 1999). Earlier work using this paradigm shows that novel word-novel object pairs can be learned quite quickly, especially by adults, eliciting visual attention (Weighall et al., 2017). Visual attention to a novel referent in earlier work is strongly associated with instances of word learning for novel word-novel object pairs (Magnuson et al., 2003) and for novel labels matched with known objects (Trueswell et al., 2013), suggesting a causal relationship between attention and novel word learning. We predicted that novel-highlighting adjectives would be associated with more visual attention to the novel image, and that attention would then increase rates of word learning.

After a short delay, participants performed a *test phase* involving yes-no judgments on label-image pairs, which informed us about how they interpreted the meaning of the novel labels. As shown in Figure 1, these test trials included the same novel and familiar images from the learning phase presented with various labels, indexing how the combination of object and label was interpreted. Conditions are labelled throughout following Figure 1; note that throughout, we use the Dutch equivalents of our example items to represent spoken stimuli (e.g. 'banaan' or 'ranaan') and use the English terms to indicate the visually presented candidate referents.

Word learning was indexed by the difference in the rate of 'yes' responses to novel labels paired with novel objects ('ranaan'+ prickly-pear) contingent on hearing the novel word 'ranaan' (condition D2 in Figure 1) or the familiar word 'banaan' (condition D1 in Figure 1) in the learning phase. Using difference scores based upon the learning phase context allows us to isolate the rate of learning given exposure to the novel word from the rate of judgments at test with no prior exposure to the novel word.

Repair was indexed by the difference in rate of 'yes' responses to novel labels presented with familiar objects regardless of learning phase context ('ranaan' + banana; conditions B1 and B2 in Figure 1) compared to novel labels presented with novel objects regardless of learning phase context ('ranaan' + prickly-pear; conditions D1 and D2 in Figure 1). This analysis focuses on responses regardless of learning phase materials because repairs can be performed at learning and at test. Here, using difference scores allowed us to isolate the rate of repair from both false alarms and novel word learning.

We also looked at whether rates of learning and repair differed based upon several additional factors. Within each experiment, we examined the roles of sentence context (novel- or familiar-highlighting), and participants' attention to images in the story phase. Following proposed mechanisms for learning and repair, we predicted that learning rates would be higher with novel-highlighting contexts, and that learning rates would be associated with more attention to the novel image and correspondingly, less attention to the familiar image. In a data set that combined both experiments, we also examined whether there were differences in rates of learning and repair based on mispronunciation size (comparing Experiment 1 and Experiment 2), and based upon individual differences in language experience. From earlier literature on lexical access in adults (Marslen-Wilson, 1993; Marslen-Wilson & Zwitserlood, 1989) and children, Von Holzen and Bergmann (2021) and on noisy channel processing (Gibson et al., 2013), we predicted that smaller mispronunciations would lead to more repair and less learning. From literature on first- and second-language learning, we predicted that extensive language experience would increase learning rates (as in e.g. Duff et al., 2015; Fernald et al., 2006; Gershkoff-Stowe & Hahn, 2007; Kaushanskaya & Marian, 2009a, 2009b; Papagno & Vallar, 1995; Perfetti et al., 2005; Van Hell & Mahn, 1997, though c.f. Muench & Creel, 2013).

2. Experiment 1

In this experiment, we contrasted rates of learning versus repair for novel words ('ranana') consistent with a several-feature mispronunciation of a target ('banana'). The experiment was preregistered on a server at the Max Planck Institute for Psycholinguistics. A copy of this preregistration and other materials are available at https://osf.io/x8qv3/.

2.1. Method

2.1.1. Participants

We recruited 22 adult participants from the Max Planck Institute for Psycholinguistics' database. Of these, two were excluded due to a computer crash, leaving 20 participants contributing to our analyses. All participants were native speakers of Dutch, with normal or corrected-to-normal vision and hearing and without a diagnosis of dyslexia. The first 11 participants were paid 8 euros for their participation, and because of a change in institute payment policy mid-study due to the COVID-19 pandemic, the remaining 11 were paid 13 euros for their participation. The study was approved by the Ethics Board of the Faculty of Social Sciences of Radboud University.

Power simulations showed that 40 participants would be sufficient to replicate a learning effect that was least 80% the size observed in an earlier study (Brehm, Kennis, Hendricks, & Bergmann, in prep, documented at https://osf.io/x8qv3/) with 80% power. We used this as an upper bound for a sequential stopping rule using the O'Brien-Fleming spending function (DeMets & Lan, 1994, taken from Schott, Rhemtulla, & Byers-Heinlein, 2019). Analyses were planned after testing 20, 30, and 40 participants with an alpha correction; this allowed us to stop data collection early at 20 participants after observing substantial evidence for the critical learning effect. We deemed this a sensible strategy because of pandemic-related participant recruitment difficulty.

2.1.2. Materials

Learning phase: In the learning phase, critical visual stimuli were comprised of 16 triads of object photos superimposed on cartoon backgrounds. Each background was a familiar indoor or outdoor location, such as classroom (see Figure 2), and contained a cartoon alien facing in the general direction of the objects or towards the viewer[2]; backgrounds were repeated across trials two to four times. Objects contained a familiar target (e.g. a banana) and a novel object (e.g. a prickly-pear) from the same category and a familiar distractor (e.g. Lego bricks) from a different category. Target labels were consonant-initial nouns of one to three syllables (M = 1.4) and relatively few phonetic neighbours (M = 8.5, range 1–24).

Images were selected using an online norming study presented on Qualtrics. In this experiment, 22 Dutch-speaking participants were recruited from the Max Planck for Psycholinguistics database. The experiment contained 42 trials with visual stimuli described as above. Each scene was paired with a written framing sentence highlighting the familiar or novel item ('She smelled a familiar/new fruit'). On each trial, participants selected the item they believed to be the referent of the sentence and then typed a name for each of the three object images. We selected 16 sets of images from this set under two criteria. First, the novel-framing sentence needed to elicit a selection of the novel target reliably more often than the familiar framing sentence, indicating that the set of 'novel' items was indeed more novel than the 'familiar' items. Second, the familiar target and familiar distractor needed to have 70% or higher name agreement and the novel object needed to have 10% or lower name agreement or to be identified only as a token of a basic-level term ('fruit'), indicating that our population of participants tended to have consistent labels for the items we expected them to know, and inconsistent labels for the items we expected them not to know. This meant that the familiar object was highly likely to be familiar, and the novel object was highly likely to be novel to all participants. An example appears in Figure 2, and more information about items can be found in Appendix A.

Each of the visual stimuli was presented alongside a short auditory passage. This contained a context sentence describing the agent's journey ('Ze zoekt verder onder de tafeltjes in het klaslokaal' ['She searched

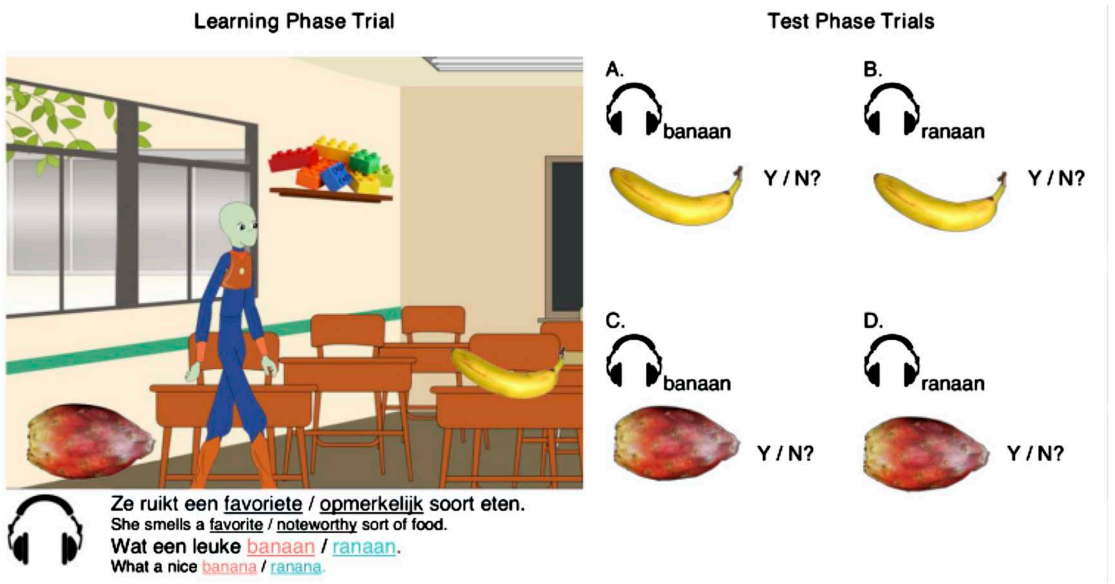

Figure 2. Example learning phase and test phase trials from Experiment 1. Per item, participants were presented with test trials of the A and D or B and C types.

further under the tables in the classroom']), followed by a sentence containing either a familiar highlighting adjective or one marking novelty and naming the semantic category of the item ('Ze ruikt een bekend/opmerkelijk soort eten' ['She smelled a familiar/remarkable kind of food']). The third sentence was the critical sentence, which labelled an object using a familiar or novel word; 'Wat een leuke banaan/ranaan' ['What a nice banana/ranana']). Auditory materials were recorded by a female native speaker of Dutch; sentences were recorded separately and cross-spliced. Four filler trials were also presented alongside the critical trials. These contained two familiar images of the same semantic category and a short audio passage.

Test phase: The test phase of the experiment consisted of images of familiar, novel, and filler items paired with familiar, novel, and distractor labels. This included 32 items testing the critical familiar and novel word pairs, and 8 items testing distractor or filler labels for a total of 40 judgments elicited from each participant. Test questions were recorded by the same speaker as the learning phase of the experiment and were worded as follows: 'Hoe heette dat ook alweer? Een [familiar/novel/distractor label] toch?' ('What was that called again? A [familiar/novel/distractor label] right?'); the first sentence was cross-spliced to three versions of the second sentence to create different versions of test stimuli.

2.1.3. Apparatus and procedure

The learning phase of the experiment was presented on a PC with an 53 cm by 30 cm flat screen monitor using Experiment Builder. Participants were seated about 75 cm from the monitor with their head in a head rest. Their eyes were tracked using an Eyelink 1000 or Eyelink 1000+. We tested each participant's dominant eye (13 R, 7 L).

The learning phase began with a 9-point calibration, followed a narrative introduction and three blocks of passive learning trials interspersed with narrative elements. These told the story of the alien landing on Earth and looking for her friend and some spaceship fuel. These narrative elements were created for an earlier child-friendly version of this study and were re-used here because we thought they might boost engagement. We selected an alien who seems familiar with Earth and has an Earth friend as the protagonist to allow participants to rely somewhat, but not entirely, on their own experience to determine what should be familiar or novel to the alien.

Learning trials were composed of scenes and sound files as described in the *Learning Phase* section above. Sound files all began with a silent interval (up to 2022 ms) so that each recording was 7400 ms long, and all ended with 800 ms of silence while the display remained on the screen. Drift checks were performed only between blocks.

After completing the learning phase of the experiment, participants performed the Dutch Peabody Picture Vocabulary Test (PPVT, Schlichting, 2005). This took 10-15 minutes. No items from the PPVT were used in the experiment– none of the PPVT items repeated any of the familiar, novel, or distractor items.

Next, participants performed the test phase of the experiment. This was presented using Presentation software and participants responded with an analogue button box. The test phase consisted of a short introduction followed by instructions, three example trials (two 'yes' and one 'no' answers), and then three blocks of critical test trials, as described in the *Test Phase* section above. There were also two transition sentences randomly inserted within each block and one transition sentence presented between each block. These contained an image of the alien from the test phase scenes and an encouraging audio clip from the same speaker (e.g. 'Oh, dat ding!' ['Oh, that thing!']). Participants responded at their own pace.

Finally, participants completed a questionnaire about their foreign language experience (see Appendix B), and were debriefed about the purpose of the experiment.

2.1.4. Design and data analysis

In the learning phase, the factors Learning Context (familiar-highlighting, novel-highlighting) and Learning Label (familiar word, novel word) were fully crossed. One version of each item was presented to each participant and the four versions of each item were assigned to four experimental lists with a Latin square. Items were presented in one of two fixed orders created by first arranging trials pseudorandomly within blocks and then reversing the order for the second list. Combined, this made a total of eight experimental lists. Six of these lists were presented to two participants each and two were presented to three participants each.

In the test phase, the Test Image (familiar, novel), Test Label (familiar, novel, distractor), and Image-Label Match (match, mismatch) were partially crossed in two lists to create image-word pairs. Each participant heard both familiar and novel labels (e.g. 'banana', 'ranana') and saw each label paired with a different image (banana, novel fruit), with an equal number of trials testing each of the four possible pairings of label and image in each list. Four of the distractor labels were also in each list, two paired with each of the familiar and novel images. Four fillers also appeared, each paired with a matching label. The two test phase lists were presented an equal number of times across participants, and were presented either once ($N = 12$ lists) or twice ($N = 4$ lists) with each learning phase list.

The first analysis examined fixations to the familiar and novel images during the learning phase of the study. This model predicted the proportion of fixations to areas of interest around the familiar and novel images in two one-second time windows (one starting one second before noun onset and ending one millisecond before noun onset, and one starting at noun onset and ending 999 ms after noun onset). Predictors in this model included the Learning Context in the learning phase (familiar-highlighting, novel-highlighting), Time Window (before noun onset, after noun onset), the label used in the learning phase (Learning Label: familiar, novel), Interest Area (familiar, novel), and their interactions, as well as Trial Number (centered, continuous). All categorical predictors were contrast coded with contrasts of −0.5, 0.5.[3]

The second analysis was a binomial mixed model implementation of signal detection theory (see DeCarlo, 1998) predicting test phase response (Response: yes, no) in the critical test trials based on whether the image and label were matched (Image-Label Match: match, mismatch; contrasts: 0, 1)[4], the test stimulus type (Test Label: familiar, novel; contrasts: 1, 0), the label presented in the learning phase (Learning Label: familiar, mispronunciation; contrasts: -0.5,0.5), and their interactions. This contrast coding sets the novel label, novel image condition as the reference level, making the effect of learning label in the model the simple effect of training for the critical novel image-label pairings. Choosing this contrast coding scheme allows us to directly test this critical pairwise comparison and maximise statistical power. We then also tested whether Learning Context (familiar-highlighting, novel-highlighting; contrasts: 0.5, −0.5) or the proportion of fixations to the familiar and novel images per participant, per trial (Proportion Familiar Fixations, Proportion Novel Fixations) that occurred after the label's onset improved the fit of this model. We did this by adding each predictor to the test phase response model, one at a time.

For all models, random intercepts were included for Participant and Item Set (each item in all its versions). All the fixed effect predictors and Trial Number (Learning Trial in analysis 1 and Test Trial in analysis 2), were initially included as additional random slopes for each random intercept; we removed random slopes that accounted for limited variance due to non-convergence, singular fit warnings, or correlations between random terms of 0.9 or above. However, when random intercepts for either Participant or Item Set led to a singular fit warning, we kept them in the model to preserve a similar effect structure across analyses. For all linear models, we take a *t*-value of 2.00 as our threshold for significance, following a reasonably common convention of the field (see Meteyard & Davies, 2020), which means that no assumptions need to be made about model degrees of freedom. All code is available at https://osf.io/x8qv3/. All analyses were conducted in R using the package lme4.

2.2. Results

First, we examined the pattern of fixations to the familiar and novel images during the learning phase of the experiment, whether this was impacted by the Learning Context or Learning Label predictors and whether it changed through the experiment. The overall pattern is depicted in Figure 3. In general, more time was spent fixating the novel image (the prickly-pear) than the familiar image (the banana), especially at the beginning of the trial. After the onset of the critical word ('ranaan' or 'banaan'), more time was spent fixating the familiar than the novel image, consistent with the expectation from disjunctive reasoning that the familiar image better matched the familiar label than the novel one. This was especially true when the familiar word was used in the trial. In order to test these patterns statistically, we examined how the proportion of fixations to each object changed across two one-second time windows (see Table 1). The zero point, aligning trials with each other, was the onset of the critical word label.

This model disclosed a main effect of Time Window, indicating that both objects received more attention after the noun onset: Window 2 had more fixations to both objects than Window 1. Time Window interacted with Interest Area, such that there was more attention to the novel object (the prickly-pear; blue in Figure 3) before noun onset and more attention to the familiar object (the banana; red in Figure 3) after noun onset. Learning Context also interacted with Interest Area. In both time windows, more attention was directed to the novel object (the prickly-pear; blue in Figure 3) in the novel-highlighting context (right panel of Figure 3), while more attention was directed to the familiar object (the banana; red in Figure 3) in the familiar-highlighting context (left panel of Figure 3). There was also an interaction between Interest Area and Learning Label, and between Time Window, Interest Area, and Learning Label. These effects mean that collapsing across time windows, more attention was directed to the novel object (the prickly-pear; blue in Figure 3) than the familiar object (the banana; red in Figure 3) when the Learning Label was the novel word ('ranaan'; dashed lines in Figure 3), and vice-versa, and that the difference was largest in the time window starting at noun onset (Window 2).

Next, we examined judgments in the test phase of the experiment. As shown in Figure 4, when the learning label was familiar ('banaan'; top row in Figure 4), rates of match responses at test were high for familiar images (a banana) paired with familiar labels ('banaan', A1: 96%), while rates of match responses at test were low for all other conditions (B1, C1, D1, between 13% and 8%). In contrast, when the learning label was novel ('ranaan'; bottom row in Figure 4), rates of match responses for the familiar image (a banana) were high for either test label ('banaan', A2: 66%, 'ranaan', B2: 73%). The novel image (a prickly-pear) was not judged to match well with familiar label ('banaan', C2: 7%) but judged to match moderately well with the novel label ('ranaan', D2: 36%).

Across the board, these data therefore show a repair rate of 5% (B1-D1) in the familiar learning label condition and 36% (B2-D2) in the novel learning label condition (comparing novel labels paired with novel, vs familiar images). They also show a learning rate of 29% (for novel words paired with novel images in the novel learning label condition compared to the known learning label baseline; D2-D1).

The effects described above were confirmed in a mixed-effect model (see Table 2). As described above, this model was contrast coded to test the critical simple effect of training on novel-label, novel-image trials. The significant intercept term corresponds to a log odds of match responses in the novel label – novel image ('ranaan' + prickly-pear) conditions that is reliably below zero, indicating a bias toward 'no' responses in conditions D1 and D2. The simple effect of Image-Label Match means that for the novel label test trials ('ranaan'), the familiar test image (the banana) elicited more match responses than the novel test image (the prickly-pear). In other words, the repair effect was reliably larger in conditions B1 and B2 than predicted from the response baseline in conditions D1 and D2. The simple effect of Test Label means that for trials where the image and label matched, the familiar label ('banaan', A1 and A2) elicited more match responses than the novel label ('ranaan', D1 and D2). This indicates that the novel words were considered to be less good matches to their image than the familiar words were. The main effect of Learning Label means that novel label- novel image test items ('ranaan' + prickly-pear) presented with the novel label at learning ('ranaan', D2) elicited more match responses than those presented with the familiar label at learning ('banaan', D1). This is the critical learning effect. One interaction was also reliable: Test Label interacted with Learning Label, such that novel label-novel image test items ('ranaan' + prickly-pear) had more match responses when the learning label was novel ('ranaan', D2) than when it was familiar ('banaan', D1), and familiar label, familiar image test items ('banaan' + banana) had more match responses when the learning label was familiar ('banaan', A1) than when it was novel ('ranaan', A2). This represents a general effect based on training, such that presenting novel labels in the learning phase decreased the goodness of familiar word-familiar item pairings.

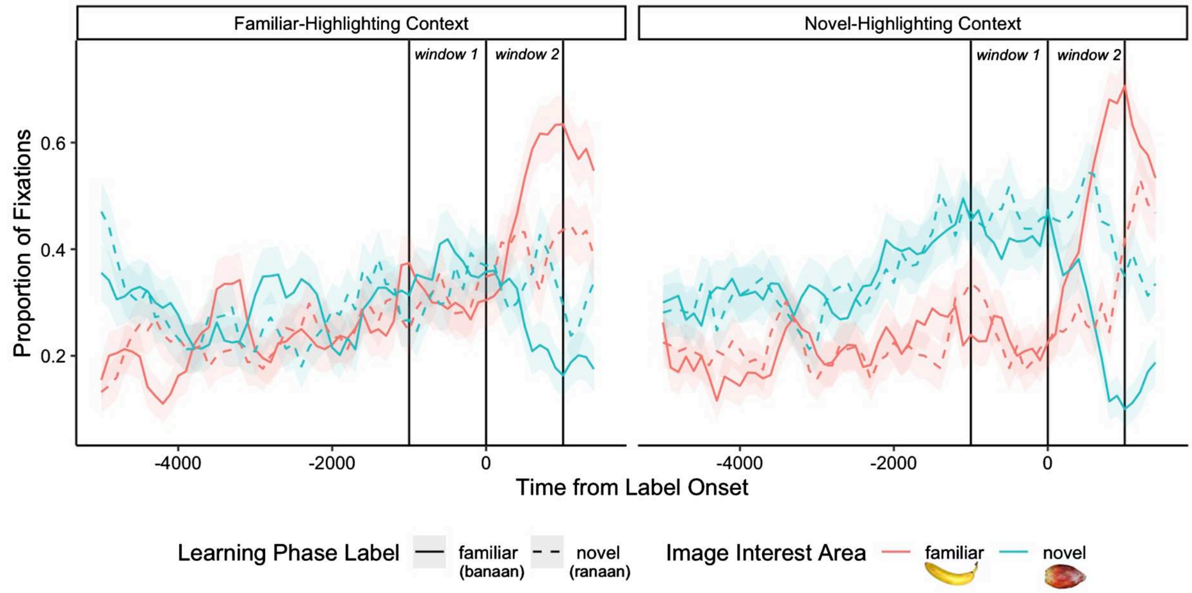

Figure 3. Proportion of fixations across trials to image interest areas during the learning phase of Experiment 1, split by Learning Context. Points reflect averages at 100 millisecond intervals, with confidence bands representing a 95% CI across observations by subject and item. Zero reflects the onset of the critical word: window 1 tests pattern of attention immediately before the critical word, and window 2 tests patterns of attention linked to the critical word. Dashed lines reflect trials where the novel word was used and solid lines reflect trials where the familiar word was used.

As shown in Table 2, when added to this model along with all interactions, Learning Context did not improve model fit, despite its association with attention to the novel image in the learning phase of the experiment. Proportion Familiar Fixations in the learning phase (e.g. average of red lines in Figure 3 per participant) also did not improve model fit, though Proportion Novel Fixations (e.g. average of blue lines in Figure 3 per participant) did. In this best-fitting model, there was a simple effect of Proportion Novel Fixations ($\beta = 2.33$, SE $= 0.98$, z-value $= 2.37$, p-value $= 0.02$), such that in the novel label- novel image reference condition ('ranaan' + prickly-pear), more fixations to the novel image (the prickly-pear) were associated with more match responses. There was also an interaction between Proportion Novel Fixations and the Test Label predictor ($\beta = -3.33$, SE $= 1.52$, z-value $= -2.20$, p-value $= 0.03$), such that in the familiar label- familiar image condition ('banaan' + banana), more fixations to the novel image (the prickly-pear) were associated with fewer match responses. In this model, the Learning Label effect was also no longer reliable, indicating that the same variance might be better captured instead by the pattern of fixations to the novel image. All other effects were comparable to what is presented in the top of Table 2. Combined, these findings show that while the adjective context in the learning phase had no direct impact on later judgments, it did impact how attention was directed to the novel image, and this attention then impacted rates of learning and repair. The implication is that while repair and learning might trade off as useful strategies, individuals may not be

Table 1. Results from linear mixed models of proportion of fixations to the familiar and novel images in pre- and post-noun onset time windows from Experiment 1.

Random effects	*SD*	*Correlation*	
Participant	0.03		
Item Set	0.03		
Interest Area x Item Set	0.20	−0.03	
Residual	0.33		
Fixed effects	*Estimate*	*SE*	*t-value*
Intercept	0.36	0.01	27.07
Interest Area	0.03	0.05	0.58
Time Window	0.04	0.02	2.31
Learning Context	0.01	0.02	0.30
Learning Label	−0.01	0.02	−0.36
Interest Area x Time Window	−0.20	0.04	−5.61
Interest Area x Learning Context	0.16	0.04	4.25
Interest Area x Learning Label	0.14	0.04	3.85
Time Window x Learning Context	−0.02	0.04	−0.47
Time Window x Learning Label	−0.02	0.04	−0.57
Learning Context x Learning Label	0.02	0.04	0.49
Interest Area x Time Window x Learning Context	−0.04	0.07	−0.61
Interest Area x Time Window x Learning Label	0.30	0.07	4.16
Interest Area x Learning Context x Learning Label	0.13	0.07	1.83
Time Window x Learning Context x Learning Label	−0.04	0.07	−0.50
Interest Area x Time Window x Learning Context x Learning Label	0.15	0.15	1.01
Trial Order	0.00	0.00	1.16

Notes: Model formula was: lmer(Proportion of Fixations ~ Interest Area * Time Window * Learning Context * Learning Label + Trial Number + (1 | Participant) + (1 + Interest Area | Item)).

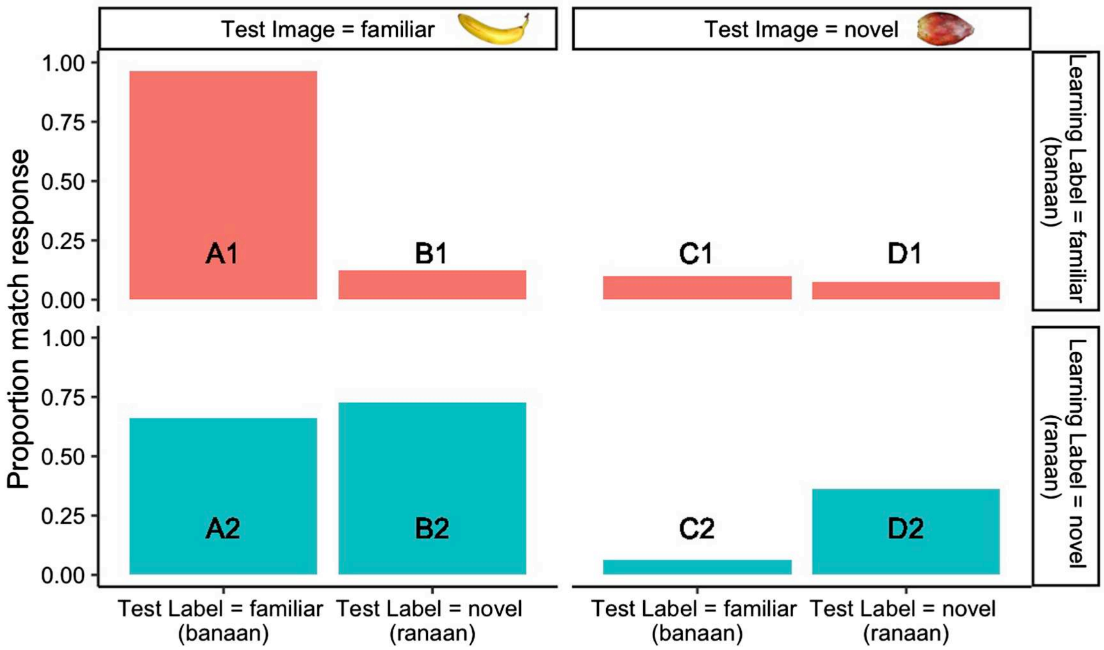

Figure 4. Proportion of match responses in critical trials by Learning Label, Test Label, and Test Image in Experiment 1.

freely choosing between these strategies while listening. We address this further in the General Discussion.

2.3. Discussion

In Experiment 1, we showed that when listeners were presented with novel words that might be interpreted as multi-feature mispronunciations of a target or as novel labels for a novel image, there was evidence for both learning and repair. Reflecting learning rates, there was an increase of about 29% of novel word-novel label match judgments when the novel label was presented in the learning phase of the experiment compared to when the known label was presented. This moderate effect size is consistent with earlier work (e.g. Coutanche & Thompson-Schill, 2014; Halberda, 2006; Storkel et al., 2006; Trueswell et al., 2013). Reflecting repair rates, there was a rate of 20% of familiar word-novel label match judgments, which was reliably larger than the false-alarm baseline. This is also consistent with earlier work (e.g. Christianson et al., 2001; Ferreira, 2003; Gibson et al., 2013; Levy, 2008; Levy et al., 2009). Repair rates were highest (36%) when the novel word was also presented in the learning phase of the experiment. This suggests that listeners in this experiment adapted to what they perceived as the types of errors produced in the study (see e.g. Ryskin et al., 2018), and suggests possible evidence for speaker-specific learning (see e.g. Palmeri et al., 1993 for a similar paradigm investigating memory for speaker variability).

Importantly, when independently measured, the sum of our estimates of repair rates (20%) and learning rates (29%) is less than 100%. This is important in light of the claim that repair and learning are alternative strategies. If on 20% of trials, participants perform a repair, and on 29% of trials, participants engage in learning, then on the remaining 51% of trials, participants's judgments suggest that they do not have sufficient evidence to

Table 2. Results from generalised linear mixed model of match responses for critical trials in Experiment 1.

Random effects	SD			
Participant	1.19			
Test Label x Participant	2.24			
Item Set	0.50			
Learning Label x Item Set	0.96			
Fixed effects	Est.	SE	*z*-value	*p*-value
Intercept (Novel Test Image, Novel Test Label)	−2.03	0.43	−4.77	< 0.001
Image-Label Match (Familiar Test Image, Novel Test Label)	1.35	0.37	3.65	< 0.001
Test Label (Familiar Test Image, Familiar Test Label)	4.66	0.74	6.30	< 0.001
Learning Label (Novel vs Familiar Learning Label in Novel, Novel Test trials)	2.49	0.62	3.99	< 0.001
Image-Label Match x Test Label	−7.25	0.79	−9.18	< 0.001
Image-Label Match x Learning Label	1.41	0.73	1.93	0.054
Test Label x Learning Label	−5.95	1.02	−5.84	< 0.001
Image-Label Match x Test Label x Learning Label	1.73	1.29	1.34	0.18
Moderating Effects	χ^2	DF	*p*-value	
Learning Context	4.38	8	0.82	
P Familiar	10.05	8	0.26	
P Novel	19.35	8	0.01	

Notes: P Familiar = Proportion Familiar Fixations and P Novel = Proportion Novel Fixations. Model formula was: glmer(Response ∼ 1 + Image-Label Match * Test Label * Learning Label * Moderator + (1 + Test Label | Subject) + (1 + Learning Label | Item)), with the moderators Learning Context, P Familiar and P Novel added one at a time.

engage in either strategy. We return to the question of what this means in the General Discussion.

Attention during the learning phase of the experiment provides evidence for the mechanisms that support learning. In general, attention was directed early in the sentence towards novel images. This is consistent with individuals following a disjunctive reasoning strategy (e.g. Halberda (2006) as they unpack the scene and make predictions about what will be relevant to the story. Once the novel word appeared, fixations declined to the novel image and increased to the familiar image. Adding the proportion of fixations to the novel image in the learning phase after the target word onset improved a model of the test phase data, which suggests that fixating the novel image is a signature of a non-literal processing strategy in the visual world paradigm.

Interestingly, while learning context supported attention in the learning phase of the experiment, it had no clear impact on the test phase judgments themselves. One possible explanation for this is participants' sensitivity to conflicting combinations of cues, such as the novel-highlighting context being repeatedly associated with a known word. A different study design would be necessary to test this possibility further. Regardless, we believe that the weak effect of learning context does still show the relative difficulty and rarity of learning by fast mapping, especially in ambiguous situations with no clear answer like the present experiment, and highlights how it is useful to combine eye-tracking and offline judgments in order to show a complete picture of processing and utterance interpretation (see e.g. Brehm et al., 2021 for more details).

3. Experiment 2

This experiment replaced the many-feature mispronunciations in Experiment 1 with a few-feature mispronunciation of the target (e.g. 'panaan' for 'banaaan'), but was otherwise identical. Our prediction was that the distance between familiar target and novel word would increase repair and decrease learning rates. We again examined visual attention to each image on its first presentation and image-label match judgments in the same delayed test paradigm.

3.1. Method

3.1.1. Participants

We recruited 26 adult participants from the Max Planck Institute for Psycholinguistics' database; no participants also participated in Experiment 1. Of these participants, one was excluded because they reported having dyslexia, one was excluded for running in a prior norming study of the materials, and four were excluded due to a splicing error in an audio file that was then corrected. This left 20 participants. All remaining participants were native speakers of Dutch, with normal or corrected to normal vision and hearing and without a diagnosis of dyslexia. This study was conducted in parallel with Experiment 1, and so again, the first 5 participants were paid 8 euros for their participation and the remaining 21 were paid 13 euros.

The same sample size calculations were performed as in Experiment 1. We used power simulations to plan an upper bound of 40 participants, with sequential stopping analyses planned after testing 20, 30, and 40 participants with an alpha correction following the O'Brien-Fleming spending function. We again stopped data collection at 20 participants after observing substantial evidence for the critical learning effect.

3.1.2. Materials

Materials were identical to Experiment 1, except that all mispronunciations now were relatively small, changing approximately 1 phonetic feature (see Appendix A for a full list of items) compared to approximately 3 in Experiment 1. We spliced sentences and target words together in the same manner as Experiment 1.

3.1.3. Apparatus and procedure

The procedure was identical to Experiment 1. We again tested each participant's dominant eye (14 R, 6 L).

3.1.4. Design and data analysis

The design and data analysis were identical to Experiment 1.

3.2. Results

We began again by examining the patterns of fixations during the learning phase of the experiment, depicted in Figure 5. The pattern was similar to Experiment 1: more time was spent fixating the novel image (the prickly-pear; blue in Figure 5) than the familiar image (the banana; red in Figure 5), especially at the beginning of the trial, and especially when the sentence context highlighted novelty (right panel in Figure 5). More time was also spent fixating the familiar than the novel image after the onset of the critical word (in Window 2), consistent with the disjunctive reasoning expectation that the familiar image better matched the familiar label than the novel one.

The same patterns were obtained statistically as in Experiment 1. Again, there was a main effect of Time Window, indicating that both objects received more attention after the noun onset (in Window 2, compared

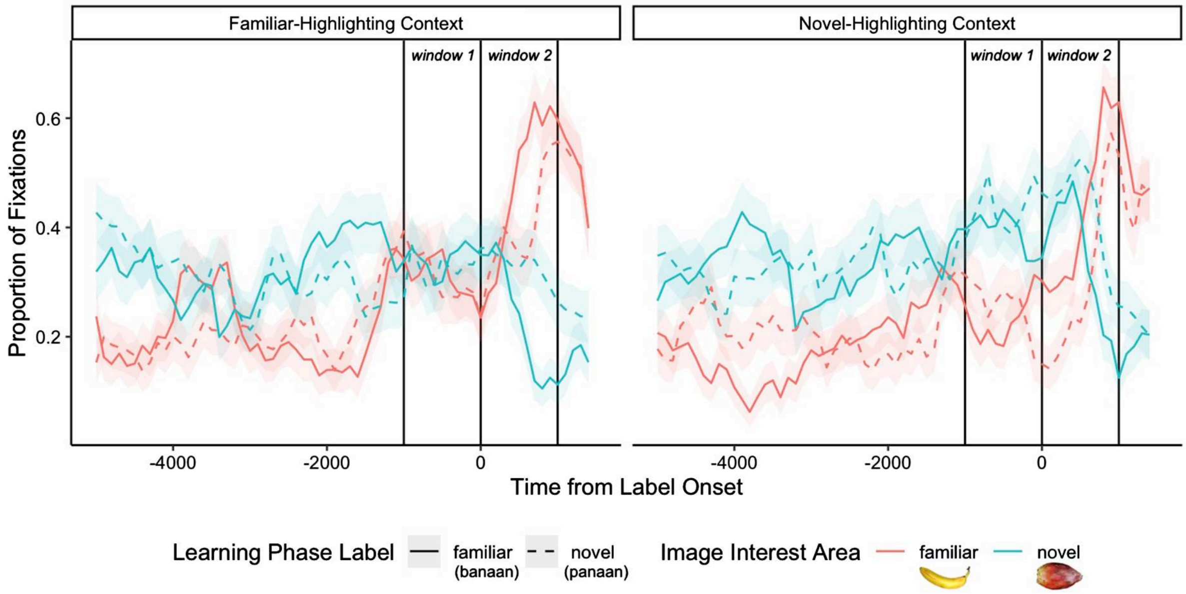

Figure 5. Proportion of fixations across trials to images during the learning phase of Experiment 2, split by Learning Context. Points reflect averages at 100 millisecond intervals, with confidence bands representing a 95% CI across observations by subject and item. Zero reflects the onset of the critical word: window 1 tests pattern of attention immediately before the critical word, and window 2 tests patterns of attention linked to the critical word. Dashed lines reflect trials where the novel word was used and solid lines reflect trials where the familiar word was used.

to Window 1). Again, Time Window interacted with Interest Area, such that there was more attention to the novel object (the prickly-pear) before noun onset in Window 1, and more attention to the familiar object (the banana) after noun onset in Window 2. Once again, Learning Context also interacted with Interest Area. In both time windows, more attention was directed to the novel object (the prickly-pear) in the novel-highlighting context, while more attention was directed to the familiar object (the banana) in the familiar-highlighting

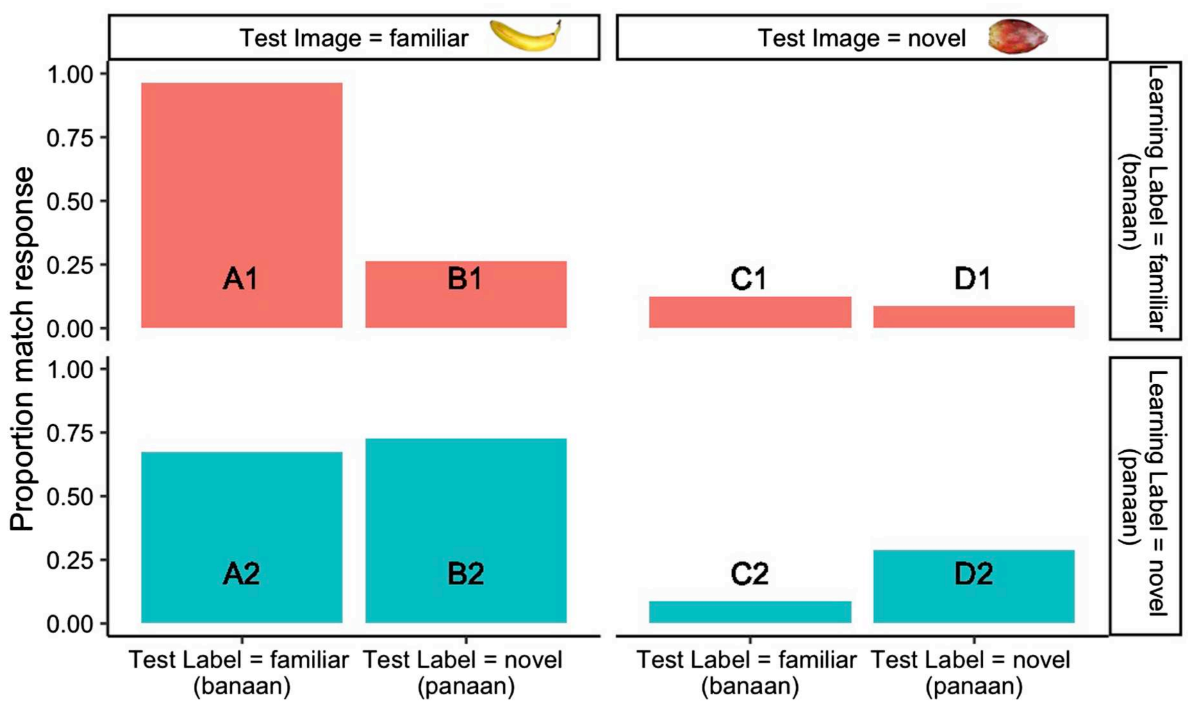

Figure 6. Proportion of match responses in critical trials by Learning Label, Test Label, and Test Image in Experiment 2.

context. Finally, there was again an interaction between Interest Area and Learning Label, and between Time Window, Interest Area, and Learning Label. Collapsing across time windows, more attention was directed to the novel object (the prickly-pear) than the familiar object (the banana) when the Learning Label was the novel word ('panaan'), and vice-versa, with the largest effect in the time window starting at noun onset (Window 2).

Next, we examined the test phase data. As shown in Figure 6, when the learning label was the familiar word ('banaan'; top row), familiar images paired with familiar labels at test ('banaan' + banana, A1) received the highest rates of match responses (96%), followed by familiar images paired with novel labels ('panaan' + banana, B1; 26%). These were followed by the novel image test trials (prickly-pear) when paired with familiar labels ('banaan', C1: 13%) and novel labels ('panaan', D1: 9%). This means compared to Experiment 1, familiar image, novel label trials elicited an additional 13% match responses.

When the learning label was the novel word ('panaan'; bottom row in Figure 6), the familiar image (banana) was judged to match fairly well with either label at test (73% for the novel label 'panaan', B2, and 68% for the familiar label 'banaan', A2). This is very similar to Experiment 1 (where familiar image match responses were 73% for the novel label, and 66% for the familiar label). There were also again moderate rates of match responses for novel label-novel image test trials ('panaan' + prickly-pear, D2: 29%), and low rates of match responses for familiar label-novel image test trials ('banaan' + prickly-pear, C2: 9%).

Combined, this shows a larger repair rate in Experiment 2 than Experiment 1: a rate of 17% in the familiar learning label condition (B1-D1), and 44% in the novel learning label condition (B2-D2), compared to rates of 5% and 36% for the same conditions in Experiment 1. Complementing these increases in repair rates, there is a smaller learning rate (D2-D1): 20% in this experiment, compared to 29% in Experiment 1.

These patterns were again confirmed in a mixed-effect model (Table 3). In this model, the intercept, corresponding to the log odds of match responses in the novel label – novel image test conditions collapsed across Learning Label, ('panaan' + prickly-pear, D1 and D2), was again reliably below zero (more 'no' than 'yes' responses). All simple effects were again reliable, and patterned as they did in Experiment 1. The simple effect of Image-Label Match meant that for the novel test label ('panaan') trials, the familiar test image (a banana, B1 and B2) elicited more match responses than the novel test image (a prickly-pear, D1 and D2); this shows that the repair effect is again reliably larger than baseline. The simple effect of Test Label meant that for images where the picture and label matched, the familiar label 'banaan', A1 and A2) again elicited more match responses than the novel label 'panaan', D1 and D2), indicating a familiar item match benefit. Finally, the main effect of Learning Label meant that for the novel label-novel image test condition ('panaan' + prickly-pear), images presented with the novel label at learning 'panaan', D2) elicited more match responses than those presented with the familiar label at learning 'banaan', D1); again, this is the critical learning effect. An interaction between Test Label and Learning Label was also again present, such that novel label- novel image test trials ('panaan' + prickly-pear) elicited more match responses when the learning label was novel ('panaan', D2) than when it was familiar ('banaan', D1), and familiar label- familiar image test trials ('banaan' + banana) elicited more match responses when the learning label was familiar ('banaan', A1) than when it was novel ('panaan', A2).

Moderating variables were again added to the model of test phase judgments, as shown in Table 4. As in Experiment 1, adding Learning Context and its interactions to this model did not improve model fit. However, both measures of the proportion of fixations to images after the onset of the critical word (Proportion Familiar Fixations, Proportion Novel Fixations– the

Table 3. Results from generalised linear mixed model of match responses for critical trials in Experiment 2.

Random effects	SD			
Participant	0.53			
Test Label x Participant	1.03			
Item Set	0.45			
Fixed effects	Est.	SE	*z*-value	*p*-value
Intercept (Novel Test Image, Novel Test Label)	−1.77	0.30	−5.92	< 0.001
Image-Label Match (Familiar Test Image, Novel Test Label)	1.73	0.31	5.60	< 0.001
Test Label (Familiar Test Image, Familiar Test Label)	3.99	0.49	8.17	< 0.001
Learning Label (Novel vs Familiar Learning Label in Novel, Novel Test trials)	1.48	0.48	3.08	< 0.01
Image-Label Match x Test Label	−6.39	0.59	−10.82	< 0.001
Image-Label Match x Learning Label	0.78	0.62	1.26	0.21
Test Label x Learning Label	−4.26	0.82	−5.20	< 0.001
Image-Label Match x Test Label x Learning Label	1.63	1.04	1.56	0.12
Moderating Effects	χ^2	DF	*p*-value	
Learning Context	12.49	8	0.13	
P Familiar	25.05	8	0.002	
P Novel	25.01	8	0.002	

Notes: P Familiar = Proportion Familiar Fixations and P Novel = Proportion Novel Fixations. Model formula was: glmer(Response ~ 1 + Image-Label Match * Test Label * Learning Label * Moderator + (1 + Test Label | Subject) + (1 | Item)), with the moderators Learning Context, P Familiar and P Novel added one at a time.

average levels of red and blue lines in Figure 5 per participant, respectively) reliably improved model fit. This shows again a link between visual attention and the choice to perform a repair or to learn a new word.

First, in the model adding Proportion Familiar Fixations, there was an interaction between Proportion Familiar Fixations and Image-Label Match ($\beta = -5.00$, SE = 1.57, z-value = −3.19, p-value < 0.01), indicating that for novel-label, familiar-image test trials ('panaan' + banana) when many fixations were made to the familiar image (the banana) in the learning phase, there was a decreased proportion of match responses in the test phase. There was also an interaction between Proportion Familiar Fixations, Image-Label Match, and Test Label ($\beta = 7.00$, SE = 2.37, z-value = 2.95, p-value < 0.01), such that while increased fixations to the familiar image (the banana) in the learning phase tended to generally decrease match responses in the test phase, this was not the case for familiar-label, familiar-image ('banaan' + banana) test trials, where increased fixations at learning increased match responses at test. Finally, there was an interaction between Proportion Familiar Fixations, Image-Label Match, and Learning Label ($\beta = 5.96$, SE = 2.18, z-value = 2.73, p-value < 0.01), such that for the matching familiar-label, familiar-image ('banaan' + banana) and novel-label, novel-image test trials ('panaan' + prickly-pear), the largest learning label effect occurred when few fixations in the learning phase were made to the familiar image, and for the mismatching familiar-label, novel-image ('banaan' + prickly-pear) and novel-label, familiar-image ('panaan' + banana) test trials, the largest learning label effect occurred when many fixations were made in the learning phase to the familiar image. In this model, the Image-Label Match effect disappeared, indicating that it was mediated by the pattern of fixations to the familiar image, but all other effects were comparable to what is presented in Table 4. Combined, the implication is that directing attention to the known image was associated with repair.

Second, in the model adding Proportion Novel Fixations, there was an interaction between Proportion Novel Fixations and Image-Label Match ($\beta = 4.98$, SE = 1.41, z-value = 3.54, p-value < 0.001), which was the inverse of the corresponding interaction in the Proportion Familiar Fixations model: for familiar-label, novel-image (prickly-pear + 'banaan') trials when many fixations were made to the novel image, there was a higher proportion of match responses. There was also an interaction between Proportion Novel Fixations, Image-Label Match, and Test Label ($\beta = -8.43$, SE = 2.17, z-value = −3.88, p-value < 0.001), indicating that for the familiar-label, familiar-image (banana + 'banaan') trials when many fixations were made to the novel image, there was a lower proportion of match responses. In this model, the effect of Test Label was no longer reliable, indicating that the Test Label effect in Table 4 was likely mediated by attention to the novel image, but all other effects were comparable to what is presented in Table 4. Combined, the implication is that directing attention to the novel image was associated with learning. This provides further evidence for bottom-up factors influencing whether one learn or performs a repair.

Table 4. Results from linear mixed models of proportion of fixations to the familiar and novel images in pre- and post-noun onset time windows from Experiment 2.

Random effects	*SD*	*Correlation*	
Participant	0.06		
Interest Area x Participant	0.14	0.08	
Item Set	0.03		
Interest Area x Item Set	0.17	−0.02	
Residual	0.31		
Fixed effects	*Estimate*	*SE*	*t-value*
Intercept	0.36	0.02	20.54
Interest Area	0.02	0.05	0.32
Time Window	0.04	0.02	2.42
Learning Context	0.01	0.02	0.73
Learning Label	0.00	0.02	0.13
Interest Area x Time Window	−0.17	0.03	−4.97
Interest Area x Learning Context	0.17	0.03	4.89
Interest Area x Learning Label	0.11	0.03	3.22
Time Window x Learning Context	0.02	0.03	0.48
Time Window x Learning Label	−0.02	0.03	−0.54
Learning Context x Learning Label	0.01	0.03	0.17
Interest Area x Time Window x Learning Context	0.04	0.07	0.55
Interest Area x Time Window x Learning Label	0.21	0.07	3.04
Interest Area x Learning Context x Learning Label	0.04	0.07	0.51
Time Window x Learning Context x Learning Label	−0.06	0.07	−0.94
Interest Area x Time Window x Learning Context x Learning Label	0.02	0.14	0.11
Trial Order	0.00	0.00	0.19

Notes: Model formula was: lmer(Proportion of Fixations ~ Interest Area * Time Window * Learning Context * Learning Label + Trial Number + (1 + Interest Area | Participant) + (1 + Interest Area | Item)).

3.3. Discussion

In Experiment 2, we presented listeners with novel words that were plausibly either small mispronunciations of a familiar target or novel labels for a novel image. Just like in Experiment 1, we showed evidence for both learning and repair. The repair rate was numerically higher than in Experiment 1: 17% in the familiar learning label condition in Experiment 2 compared to 5% in Experiment 1, and 44% in the novel learning label condition compared to 36% in Experiment 1. This suggests that, consistent with earlier work on lexical access (Marslen-Wilson, 1993; Marslen-Wilson & Welsh, 1978) and theories of non-literal processing, listeners make more repairs for more plausible errors. The learning rate was also numerically lower: 20% in Experiment 2 compared to 29% in Experiment 1. The simultaneous increase in repair rate and decrease of learning rate also provides further evidence that learning and repair are indeed opposing strategies, with repair being the default in most settings.

As in Experiment 1, attention during the learning phase of Experiment 2 was directed to images in a predictable, sensible way. Participants looked at the novel image and then after hearing the onset of the critical word, began to look at the familiar image more often. Fixations to both the novel and familiar items in the learning phase reliably predicted test phase judgments, such that fixations to the familiar image during a late time window coincided with repair, and fixations to a the novel image in the same window coincided with learning. This provides further support that attention to the familiar image is a behavioural index of performing a repair, whether this is undertaken implicitly or strategically.

4. Across-experiment comparison and individual differences

In a final analysis, the test phase judgment data from both experiments was combined to examine the impact of several additional mediating variables. We predicted that small mispronunciations would increase repair rates (e.g. Gibson et al., 2013; Marslen-Wilson, 1993; Marslen-Wilson & Welsh, 1978; Von Holzen & Bergmann, 2021), and that increased language experience would increase learning rates (as in Duff et al., 2015; Fernald et al., 2006; Gershkoff-Stowe & Hahn, 2007; Kaushanskaya & Marian, 2009a, 2009b; Papagno & Vallar, 1995; Perfetti et al., 2005; Van Hell & Mahn, 1997).

We first tested whether adding the predictor Experiment (contrast coded −0.5, 0.5) and its interactions with all factors improved model fit. This tested the role of mispronunciation size in rates of repair and learning. Despite the numerical differences between experiments, adding this factor did not improve model fit ($\chi^2(8) = 6.60$, $p = 0.58$), nor were there any additional significant effects in a model including Experiment. Because of the relatively small total sample size ($N = 40$) this analysis is likely underpowered. Regardless, it does cast some doubt on the conclusion that listeners are strongly affected by phonetic mispronunciation size in this paradigm: if mispronunciation size does have a true effect on the balance of novel word learning and repair, it is relatively small.

Centered versions of PPVT score (untransformed range: 155–195, $M = 177.5$, $SD = 9.02$), the number of language spoken by the participant (untransformed range 2–6, $M = 3.6$, $SD = 1.04$), and the self-reported score of participant's most proficient foreign language (untransformed range 2–5, $M = 3.9$, $SD = 0.92$), were also each added to the model one at a time along with their interactions. PPVT score came closest to improving the model fit, but did not do so reliably ($\chi^2(8) = 14.14$, $p = 0.08$). Similarly, neither the number of languages spoken ($\chi^2(8) = 9.90$, $p = 0.27$) nor highest foreign language proficiency ($\chi^2(8) = 9.84$, $p = 0.28$) improved model fit. Again, the relatively small total sample size suggests that these analyses are likely underpowered; however, we highlight vocabulary score in particular as a promising direction for future work on individual differences in word learning versus repair because of its general trend towards improving model fit, in line with previous work.

5. General discussion

When faced with an ambiguous novel word like the Dutch item 'ranaan', listeners can choose to interpret it as either a mispronunciation of a familiar target ('ranaan' repaired to 'banaan') or a novel label for a novel object ('ranaan' fast-mapped to a new item). By combining visual-world eye-tracking with an offline judgment paradigm, we showed clear evidence for both repair and fast-mapping interpretations of novel words that we created by large (Experiment 1) and small phonetic feature (Experiment 2) edits to the onset of a familiar target.

Across both experiments, participants performed both repair and fast-mapping. Repair was the dominant strategy: participants often judged that the novel label and known image matched, especially when the novel word was also presented in the learning phase (with a repair rate of 36% in Experiment 1 and 44% in Experiment 2). However, they also acquired the novel word as a label for a novel item fairly often (29% in Experiment 1 and 20% in Experiment 2). Importantly, because these results rely on match/mismatch judgments on separate test trials, these outcomes are independent, and any remaining percentage most likely reflects participants' preference for a third option, such as being uncertain which label is most appropriate for each item. The fact that repair and learning outcomes sum to less than 100% suggests that they are probably opposing strategies, and that participants likely maintain some uncertainty about which label is most appropriate. Theoretically, the implications are that (1) repair and learning are both strategies that participants entertain for ambiguous items, and (2) disjunctive reasoning about how objects and labels are paired together may support both repair and learning in this context.

Further support for repair and learning as competing strategies comes from investigating visual attention during the learning phase of the experiment. Learning was strongly associated with the amount of attention directed to the novel image after the onset of the target word. This shows how disjunctive reasoning affords the fast mapping of novel labels to referents and suggests that attention to images during listening is causally related to how an ambiguous word will be interpreted at a delayed test. In other words: one can predict a participant's later response in this paradigm with reasonable accuracy by measuring where they placed their visual attention on their first exposure to a new word. Being able to predict a later response from an earlier behaviour may prove useful as an online diagnostic of processing strategies in future work.

Importantly, there was also evidence for uncertainty of whether to repair or fast-map. The rates of repair and fast-mapping for novel labels summed to 65% in Experiment 1 and 64% in Experiment 2. The remaining 35% and 36% of trials, respectively, likely reflects a belief that neither answer is correct. The most interesting reason for this is that participants might decide that they have insufficient evidence to determine what the novel word maps to, and may instead maintain uncertainty about the meaning of the referent. Examining how online attention and offline judgments change with multiple presentations of a novel word (as in Trueswell et al., 2013) would provide clarity as to how these three strategies trade off, and is an intriguing question for future work.

As outlined in the Introduction, we predicted that several cues would influence whether novel words would be repaired or learned. From the non-literal processing literature, we predicted that few-feature errors would be associated with more repair. From the fast mapping literature, we predicted that novel-highlighting sentence contexts and visual attention to the novel image in the learning phase would be associated with more learning and less repair, and from the word learning literature, we predicted that language experience would increase learning and decrease repair. We discuss each of these factors in turn below.

Mispronunciation size, manipulated between experiments, led to numeric differences in the overall rate of novel word learning versus repair. However, it did not produce any reliable effect when added as a predictor in a joint analysis of both experiments. In the current study, we see onset mispronunciation size has either no effect or quite a small effect on word learning versus repair. This is contrary to what is predicted from work focussing on lexical access in simpler contexts (e.g. Krueger et al., 2018; Marslen-Wilson, 1993; Marslen-Wilson & Zwitserlood, 1989; Von Holzen & Bergmann, 2021), where mispronunciation size has a robust effect, and from the noisy channel literature, where edit distance reliably affects non-literal processing (e.g. Gibson et al., 2013, 2017; Ryskin et al., 2018). One possible reason for this is that onset errors might lead to a strong repair bias in general. Some developmental literature is consistent with this finding (Swingley, 2016), though c.f. (Krueger et al., 2018), who found children to be sensitive to plausible onset substitutions. Another possible difference lies in the modality: most noisy channel literature focuses on written errors, and speech and writing errors might be processed differently. While phonetic features are sometimes swapped in speech errors (for example, changing voicing to exchange /b/ with /p/, Fromkin, 1971; MacKay, 1970), many speech errors actually occur on the phoneme level (exchanging /b/with any other consonant, Shattuck-Hufnagel & Klatt, 1979). This could minimise the effect of feature size in listeners' expectations of spoken errors. Both claims would be testable in future work.

Learning context had a reliable impact on attention during the learning phase of the experiment: novel-highlighting contexts drew attention to the novel image and away from the familiar image. This follows from earlier work on disjunctive reasoning in fast mapping (Halberda, 2006; Trueswell et al., 2013) and provides evidence for online attention predicting later offline behaviour. Critically, learning context did not itself directly impact later judgments, though attention to the novel image in late time windows of sentences in the learning phase was strongly associated with more learning and less repair. This suggests that contextual salience might change the balance of repair and learning by modulating attention, not by promoting one strategy over another: later decisions were better predicted by what participants actually attended to than by our attempts to manipulate their attention.

The suggestion that contextual salience mainly affects attention dovetails with a separate literature on the broader effects of linguistic focus on sentence processing: focus has been proposed to increase attention to particular items which then enhances their processing and increases memory accuracy (e.g. Birch & Garnsey, 1995; Cutler & Fodor, 1979; Fraundorf et al., 2010; Sturt et al., 2004). From this, we speculate that other top-down factors such as object salience or joint attention might also support learning in the paradigm we have developed here (see e.g. Moore et al., 1999 or Yurovsky et al., 2017 for analogous findings with children). Top-down factors that affect attention or other item-level differences might be especially worthy of future study.

We observed no reliable effects of any measure of language ability, counter to what we predicted based on earlier literature (Kaushanskaya & Marian, 2009a, 2009b; Papagno & Vallar, 1995; Van Hell & Mahn, 1997). In a combined analysis of both experiments, we examined the impact of vocabulary test scores, the number of languages spoken by each participant, and the highest self-reported proficiency in any second language on learning and repair rates. No factors improved model fit. One possible reason for this null result is that all effect sizes are relatively small, requiring more participants and broader variability on each measure. In this study, all participants had relatively high vocabulary scores, and all were multilingual with most participants speaking three or more languages. This reduced range in our observed variables makes it statistically difficult to explore effects of each factor. Another possibility is that language experience is a multi-dimensional factor that sometimes leads to complex patterns of results. For example, the language learning advantage for bilinguals compared to monolinguals does not always replicate (see Muench & Creel, 2013). In addition, not all multilinguals are alike in their language habits and their degree of connection to multiple language communities. These variations in experience might matter for patterns of repair, in addition to learning: multilinguals immersed in multilingual communities make better inferences about the mental states of others (Tiv et al., 2021) and have a better ability to adapt to pronunciation differences from foreign accents (Kutlu et al., 2022). As such, it would be worthwhile to explore the role of these and other individual differences in a similar paradigm with a larger sample size, richer measures of language experience, and a sampling procedure that targets a more diverse set of participants, for example, through online testing.

The lack of effect of individual difference factors might also be explainable to some extent by properties of our particular items and how they interact with our participant sample. While we motivated our vocabulary predictions in terms of entrenchment promoting learning (e.g. Perfetti & Hart, 2002), vocabulary size also affects the structure of the mental lexicon. Larger vocabularies have more words in them, making them denser, with more close neighbours. It is known that close neighbours are particularly important drivers of competition in lexical access (Chen & Mirman, 2012). Items from dense phonological neighbourhoods also tend to be relatively harder to learn (Storkel et al., 2006), plausibly because they are confusable with a larger number of pre-existing items in the lexicon. Combined, this leads to the prediction that items from dense phonological neighbourhoods may be less learnable, particularly by individuals with large vocabularies. The items selected for our study had to have relatively few phonological neighbours in order for us to be able to do the three-feature change manipulation we did in Experiment 2, and all items were short and relatively frequent. It might be especially fruitful to investigate the role of individual differences in learning versus repair within items in dense and sparse phonological neighbourhoods.

We now turn to some possible extensions of the paradigm and future directions. First, we note that the fact that the presentation of materials in a story context diverges from the typical visual world eye-tracking paradigm. The pragmatics of the experimental context are therefore important to consider here. We designed the experiment to use a narrative context because we had originally developed the materials for a study with children and thought that reusing them here in the same form would make the study fun but still work as an implementation of the visual-world paradigm. Indeed, participants did tend to report that they found the experiment enjoyable in our debriefing.

However, the narrative context might itself change overall response and eye fixation patterns (see Luke & Asplund, 2018 for evidence for reduced rates of prediction-related fixations in a similar story reading task). As pointed out by a reviewer, the narrative context might also enhance demand characteristics such that participants' buy-in to a narrative make them try harder to behave as they are 'supposed to'. We cannot rule out this possibility, but we do note that in other contexts, studies designed to be more fun and more ecologically valid still elicit similar data patterns as observed in more typical lab studies (see e.g. Long et al., 2023; Speed et al., 2017). Similarly, we used an alien as a protagonist in the story because we had hoped this would be supportive of both repair (because she seems familiar with Earth objects) and learning (because she is an alien). However, as a reviewer notes, it is possible that use of an alien character could instead have prompted participants to explicitly pick a strategy to use, meaning that our attention measures index explicit strategy use rather than implicit, processing. As such, the current work– much like a lot of other work in psycholinguistics– might be best viewed as representing processing in one particular context, rather than processing in general.

We therefore take the present work as a starting point for future research on learning versus repair, where one can begin to disentangle strategy use by taking multiple measurements of each image-label combination. In future work, it would be useful to directly assess the role of demand characteristics or strategy use in participants' interpretation of ambiguous words. To test this question, future work might manipulate participants' expectations within the experiment by using a less narrative paradigm, by changing the story presented in the narrative, or by using a different protagonist. This could inform our understanding of the role of context and prior beliefs in sentence interpretation, and would further enhance our understanding of task effects in the visual world paradigm compared to more naturalistic language situations.

Finally, we return to an interesting but unexpected finding: the repair effect in this experiment was largest when the novel word was re-presented in the learning phase of the experiment, which is an important finding to replicate and extend. We suggest that this is evidence for within-experiment learning. Participants are exquisitely sensitive to the expectations of an experiment, adapting to the presence of novel structures (Fine et al., 2013) and to the base rates of various types of errors (Ryskin et al., 2018). Similarly, it is also possible that any learning observed (whether of labels or within-experiment contingencies) is potentially speaker-specific, tied to the repetition of the speaker producing the items (e.g. Brown-Schmidt, 2009; Eisner & McQueen, 2005; Palmeri et al., 1993 for evidence for speaker-specific learning at various levels of linguistic representation). Future work might change expectations within the study by presenting different filler trials, with errors or other source of variation, and could test whether this learning was speaker-specific by having learning and test items produced by different speakers.

6. Conclusion

When is a 'ranana' a 'banana'? Fairly often, given the right context. When listeners were faced with an ambiguous word that could plausibly be either a speech error or the label for a novel referent, they were more likely to perform a repair. However, novel word learning also frequently occurred, and was supported by visual attention to a novel image when hearing the novel word for the first time. This underscores that repair and learning are valid but opposing strategies, and shows how listening is an active process requiring developing and revising predictions about what should come next and about what is being referred to now.

Notes

1. Feature are the basic unit of sounds, including sound class (glides/other consonants), voicing (voiced/unvoiced), place of articulation (location of articulators), and manner of articulation (air flow shape). /r/ differs from /b/ in voicing, place, and manner, while /p/ differs from /b/ only in voicing.
2. Based on a suggestion from a reviewer, we tested if the alien's eye gaze, coded as looking towards or away from the novel item, impacted performance in an aggregate data set from both experiments. While there were more match responses for trials where the alien was looking away from the novel item (41% to 35%), adding this factor did not improve model fit and was not associated with any significant effects in the model. We therefore conclude that this cue was not an important driver of our results.
3. In our preregistration, we planned to use polynomial models with time as a continuous factor to examine these data. These proved hard to interpret, especially in the face of complex interactions, and on a reviewer's suggestion, we instead present a simpler windowed analysis.
4. This counts the novel image as 'matching' the novel label, regardless as to whether the two appeared together for the participant in the study's learning phase.

Acknowledgments

We would like to thank Caroline Rowland and Antje Meyer for helpful comments and for their support with the project.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by the MPI [Levelt Innovation Award] for Psycholinguistics to Laurel Brehm and Christina Bergmann.

Author contributions

LB: Stimulus and protocol development, funding acquisition, data collection, analysis, writing, editing. CB: Stimulus and protocol development, funding acquisition, data collection, writing, editing. NK: Stimulus development, data collection, writing, editing.

Reproducibility

All code, materials, and data required to reproduce this research are publicly available and documented (https://osf.io/x8qv3/; Bergmann et al. (2023)).

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference.

Cognition, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Bailey, K. G., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language*, *49*(2), 183–200. https://doi.org/10.1016/S0749-596X(03)00027-5

Bergmann, C., Brehm, L., & Kennis, N. (2023). Learning in a noisy channel. osf.io/x8qv3

Birch, S. L., & Garnsey, S. M. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language*, *34*(2), 232–267. https://doi.org/10.1006/jmla.1995.1011

Brehm, L., Jackson, C. N., & Miller, K. L. (2019). Speaker-specific processing of anomalous utterances. *Quarterly Journal of Experimental Psychology*, *72*(4), 764–778. https://doi.org/10.1177/1747021818765547

Brehm, L., Jackson, C. N., & Miller, K. L. (2021). Probabilistic online processing of sentence anomalies. *Language, Cognition and Neuroscience*, *36*(8), 959–983. https://doi.org/10.1080/23273798.2021.1900579

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*(2), 171–190. https://doi.org/10.1016/j.jml.2009.04.003

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, *15*, 17–29.

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, *119*(2), 417–430. https://doi.org/10.1037/a0027175

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Christianson, K., & Luke, S. G. (2011). Context strengthens initial misinterpretations of text. *Scientific Studies of Reading*, *15*(2), 136–166. https://doi.org/10.1080/10888431003636787

Cole, R. A. (1973). Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics*, *13*(1), 153–156. https://doi.org/10.3758/BF03207252

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Fast mapping rapidly integrates information into existing memory networks. *Journal of Experimental Psychology: General*, *143*(6), 2296–2303. https://doi.org/10.1037/xge0000020

Coutanche, M. N., & Thompson-Schill, S. L. (2015). Rapid consolidation of new knowledge in adulthood via fast mapping. *Trends in Cognitive Sciences*, *19*(9), 486–488. https://doi.org/10.1016/j.tics.2015.06.001

Creel, S. C. (2012). Phonological similarity and mutual exclusivity: On-line recognition of atypical pronunciations in 3–5-year-olds. *Developmental Science*, *15*(5), 697–713. https://doi.org/10.1111/desc.2012.15.issue-5

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*(1), 49–59. https://doi.org/10.1016/0010-0277(79)90010-6

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, *3*(2), 186–205. https://doi.org/10.1037/1082-989X.3.2.186

Duff, D., Tomblin, J. B., & Catts, H. (2015). The influence of reading on vocabulary growth: A case for a matthew effect. *Journal of Speech, Language, and Hearing Research*, *58*(3), 853–864. https://doi.org/10.1044/2015_JSLHR-L-13-0310

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*(2), 224–238. https://doi.org/10.3758/BF03206487

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*(1), 98–116. https://doi.org/10.1037/0012-1649.42.1.98

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS One*, *8*(10), e77661. https://doi.org/10.1371/journal.pone.0077661

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language*, *63*(3), 367–386. https://doi.org/10.1016/j.jml.2010.06.004

Frazier, L., & Clifton Jr, C. (2011). Quantifiers undone: Reversing predictable speech errors in comprehension. *Language*, *87*(1), 158–171. https://doi.org/10.1353/lan.2011.0024

Frazier, L., & Clifton Jr, C. (2015). Without his shirt off he saved the child from almost drowning: Interpreting an uncertain input. *Language, Cognition and Neuroscience*, *30*(6), 635–647. https://doi.org/10.1080/23273798.2014.995109

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*, 27–52. https://doi.org/10.2307/412187

Gershkoff-Stowe, L., & Hahn, E. R. (2007). Fast mapping skills in the developing lexicon. *Journal of Speech, Language, and Hearing Research*, *50*(3), 682–697. https://doi.org/10.1044/1092-4388(2007/048)

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056. https://doi.org/10.1073/pnas.1216438110

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't underestimate the benefits of being misunderstood. *Psychological Science*, *28*(6), 703–712. https://doi.org/10.1177/0956797617690277

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, *53*(4), 310–344. https://doi.org/10.1016/j.cogpsych.2006.04.003

Heibeck, T. H., & Markman, E. M. (1987). Word learning in children: An examination of fast mapping. *Child Development*, *58*(4), 1021–1034. https://doi.org/10.2307/1130543

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, *69*(5), 1013–1040. https://doi.org/10.1080/17470218.2015.1053951

Kaushanskaya, M., & Marian, V. (2009a). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, *16*(4), 705–710. https://doi.org/10.3758/PBR.16.4.705

Kaushanskaya, M., & Marian, V. (2009b). Bilingualism reduces native-language interference during novel-word learning.

Journal of Experimental Psychology: Learning, Memory, and Cognition, *35*(3), 829–835.

Keshev, M., & Meltzer-Asscher, A. (2021). Noisy is better than rare: Comprehenders compromise subject-verb agreement to form more probable linguistic structures. *Cognitive Psychology*, *124*, 101359. https://doi.org/10.1016/j.cogpsych.2020.101359

Krueger, B. I., Storkel, H. L., & Minai, U. (2018). The influence of misarticulations on children's word identification and processing. *Journal of Speech, Language, and Hearing Research*, *61*(4), 820–836. https://doi.org/10.1044/2017_JSLHR-S-16-0379

Kukona, A., Braze, D., Johns, C. L., Mencl, W. E., Van Dyke, J. A., Magnuson, J. S., & Tabor, W. (2016). The real-time prediction and inhibition of linguistic outcomes: Effects of language and literacy skill. *Acta Psychologica*, *171*, 72–84. https://doi.org/10.1016/j.actpsy.2016.09.009

Kutlu, E., Tiv, M., Wulff, S., & Titone, D. (2022). Does race impact speech perception? An account of accented speech in two different multilingual locales. *Cognitive Research: Principles and Implications*, *7*(1), 1–16.

Levy, R. (2008). A noisy-channel model of human sentence comprehension under uncertain input. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (pp. 234–243). Association for Computational Linguistics.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), 21086–21090. https://doi.org/10.1073/pnas.0907664106

Long, B., Simson, J., Buxó-Lugo, A., Watson, D. G., & Mehr, S. A. (2023). How games can make behavioural science better. *Nature*, *613*(7944), 433–436. https://doi.org/10.1038/d41586-023-00065-6

Lowder, M. W., & Ferreira, F. (2019). I see what you meant to say: Anticipating speech errors during online sentence processing. *Journal of Experimental Psychology: General*, *148*(10), 1849–1858. https://doi.org/10.1037/xge0000544

Luke, S. G., & Asplund, A. (2018). Prereaders' eye movements during shared storybook reading are language-mediated but not predictive. *Visual Cognition*, *26*(5), 351–365. https://doi.org/10.1080/13506285.2018.1452323

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, *8*(3), 323–350. https://doi.org/10.1016/0028-3932(70)90078-3

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*(2), 202–227. https://doi.org/10.1037/0096-3445.132.2.202

Marslen-Wilson, W. (1993). Issues of process and representation in lexical access. In G. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 187–210). Taylor & Francis.

Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29–63. https://doi.org/10.1016/0010-0285(78)90018-X

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 576–585.

Meteyard, L., & Davies, R. A. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*, 104092. https://doi.org/10.1016/j.jml.2020.104092

Moore, C., Angelopoulos, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology*, *35*(1), 60–68. https://doi.org/10.1037/0012-1649.35.1.60

Muench, K. L., & Creel, S. C. (2013). Gradient phonological inconsistency affects vocabulary learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1585–1600.

Newport, E. L., Bavelier, D., & Neville, H. J. (2001). Critical thinking about critical periods: Perspectives on a critical period for language acquisition. In E. Dupoux (Ed.), *Language, brain and cognitive development: Essays in honor of Jacques Mehler* (pp. 481–502). MIT Press.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395. https://doi.org/10.1037/0033-295X.115.2.357

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309–328.

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *The Quarterly Journal of Experimental Psychology Section A*, *48*(1), 98–107. https://doi.org/10.1080/14640749508401378

Patson, N. D., & Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *Quarterly Journal of Experimental Psychology*, *69*(5), 950–971. https://doi.org/10.1080/17470218.2014.992445

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of Functional Literacy*, *11*, 67–86.

Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1281–1292.

Ryskin, R., Futrell, R., Kiran, S., & Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition*, *181*, 141–150. https://doi.org/10.1016/j.cognition.2018.08.018

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, *6*(9), 382–386. https://doi.org/10.1016/S1364-6613(02)01958-7

Schlichting, L. (2005). *Peabody picture vocabulary test, dutch version*. Pearson Assessment and Information BV.

Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, *18*(1), 41–55. https://doi.org/10.1016/S0022-5371(79)90554-1

Speed, L. J., Wnuk, E., & Majid, A. (2017). Studying psycholinguistics out of the lab. In A. de Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics and the neurobiology of language: A practical guide* (pp. 190–207). Wiley.

Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology*, *109*(1), 132–140. https://doi.org/10.1016/j.jecp.2010.10.013

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research*, *49*(6), 1175–1192. https://doi.org/10.1044/1092-4388(2006/085)

Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review*, *11*(5), 882–888. https://doi.org/10.3758/BF03196716

Swingley, D. (2016). Two-year-olds interpret novel phonological neighbors as familiar words. *Developmental Psychology*, *52*(7), 1011–1023. https://doi.org/10.1037/dev0000114

Tiv, M., O'Regan, E., & Titone, D. (2021). In a bilingual state of mind: Investigating the continuous relationship between bilingual language experience and mentalizing. *Bilingualism: Language and Cognition*, *24*(5), 918–931. https://doi.org/10.1017/S1366728921000225

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–156. https://doi.org/10.1016/j.cogpsych.2012.10.001

Van Hell, J. G., & Mahn, A. C. (1997). Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. *Language Learning*, *47*(3), 507–546. https://doi.org/10.1111/lang.1997.47.issue-3

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging*, *18*(2), 332–339. https://doi.org/10.1037/0882-7974.18.2.332

Von Holzen, K., & Bergmann, C. (2021). The development of infants' responses to mispronunciations: A meta-analysis. *Developmental Psychology*, *57*(1), 1–18. https://doi.org/10.1037/dev0001141

Weighall, A., Henderson, L. M.., Barr, D., Cairney, S. A., & Gaskell, M. G. (2017). Eye-tracking the time-course of novel word learning and lexical competition in adults and children. *Brain and Language*, *167*, 13–27. https://doi.org/10.1016/j.bandl.2016.07.010

Yurovsky, D., Case, S., & Frank, M. C. (2017). Preschoolers flexibly adapt to linguistic input in a noisy channel. *Psychological Science*, *28*(1), 132–140. https://doi.org/10.1177/0956797616668557